Interim Report to the
National Aeronautics and Space Administration
Grant NsG 81-60

DENDRAL

A SYSTEM FOR COMPUTER CONSTRUCTION, ENUMERATION AND NOTATION OF ORGANIC MOLECULES AS TREE STRUCTURES AND CYCLIC GRAPHS

Submitted by

Joshua Lederberg
Professor of Genetics
School of Medicine
Stanford University
Palo Alto, California

Studies related to this report have been supported by research grants from the National Aeronautics and Space Administration (NsG 81-60), and from the Advanced Research Projects Agency of the Office of the Secretary of Defense (SD-183).

Part III. March 13, 1968

DRAFT VERSION

Provisional version of DENDRAL. Part III

Complete Chemical Graphs; Embedding Rings in Trees.

.01 This preliminary version is released at this time in order to facilitate the programming of DENDRAL for ringed structures, in particular the embedding of rings in trees.

The external notation is subject to further revision. A more compact one can easily be devised that also provides for implicit data and merges some of the lists. For canonical ordering, it may be necessary to expand such a compact formula back to the explicity LISP lists of the generator programs, for example the edge count list, path modifier list, and vertex modifier list could be unified

For the purposes of this report, we focus on the internal notation and the combinatorial and automorphism problems that have to be faced by the generator program. Formulas will be lists that can be read by a LISP language processor. Note that commas are redundant. "." is, however, in no case the dot of a LISP dotted pair, and this character should be translated at input-output.

This change of emphasis modifies some of the hierarchical choices, but none of the fundamental ideas developed in DENDRAL I (acyclic molecules) and DENDRAL II (general survey of regular cyclic trivalent graphs).

INDEX

Section 3.1 is a glossary that should be consulted together with Table 3.1 as an authoritative definition of DENDRAL valuation.

Section 3.2 elaborates the completion of ring definitions as mappings on the nodes and edges of one of the VG's listed in Dendral II.

Section 3.3 specifies how rings are embedded in trees to complete the notation outlined in Dendral I.

Section 3.4 deals with some problems of chirality

Sections 3.5, ff, will be completed at a later date. ----

They will cover such topics as

- abbreviated notations for human interaction
- Permanent labels and special mappings for frequently encountered rings
- Special treatment of aromaticity
- Efficient algorithism for dealing with symmetries
- Optional hierarchies of DENDRAL classification
- Heuristic questions of plausible structures
- Rings with tetravalent (spiro) vertices

3,200 CANONICAL FORMS

In general, the canonical choice among automorphisms is the lowest valued vector description of the structure, evaluated cell by cell. It is important to follow the standard hierarchy of choice, as given in Table 3.1. For example, the pendant radicals are listed first, in ascending DENDRAL order, before the lowest locant vector is selected.

This convention is consistent with DENDRAL I and facilitates interfacing the computer programs for linear and cyclic DENDRAL.

RING. A molecule which is a pure cyclic structure, with no appended twigs.

RINGED COMPOUND. A tree, possibily degenerate, in which a ring is embedded.

R-TREE. A representation in which any ring is represented as a superatom.

SUPERATOM. A node representing a previously defined complex of atoms, treated as an item by the generator

VERTEX GROUP, VG, CUBICAL GRAPH. An abstract, cubical (trivalent) graph, summarized in DENDRAL 2, on whose nodes and edges are mapped the vertex atoms and connecting linear paths of the ring.

COMPOSITION. List of atoms comprising a molecule, e.g. (C3N1O2U1) or (C3H7NO2).

R-COMPOSITION. A composition in which ring atoms are removed and allocated to ring identifiers. In effect, the composition of an R-tree.

LOCANT. An ordinal number specifying an atom or a bond in a molecule for the purpose of attaching a radical. or substituting a heteroatom

CHIRALITY. Structural information not given by the topological connectivity. This usually concerns the orientation in space of tetrahedral carbon atoms with asymmetrically attached radicals.

GENERATOR. A program to generate a complete but irredundant list of isomers of a given composition.

ORTHOMESH. A particular ring system presented as if all heteroatoms are replaced by C atoms. It is specified by a VG plus an RCL in canonical order.

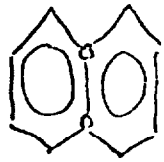
EDGE COUNT LIST, ECL A mapping of path lengths on to the edges of a VG.

.21 RING DEFINITION

A ring will customarily be defined (as a superatom) at the head of a formula. That is to say, the generator will allocate progressively larger numbers of atoms to one or more rings, then build all possible R-trees compatible with that R-composition. This canonical sequence can, of course, be rearranged for heuristic purposes.

Each predefined ring will be given an arbitrary temporary label, x2, etc. Some rings may be permanently defined, since they occur so frequently. The most prevalent example is the aromatic, benzene ring which is defined under the name Z6 as

Z6* benzene



ZN* naphthalene

Z6* = ((o) 6*()()) ; Note also

ZN* = ((2Λ) (o*, 4*, 4*)()())

As illustrated by this example, a ring definition will have the form

orthomesh

VG ECL VH PH

xn = ((vertex group) (edge count list)(vertex modifiers) (path modifiers)),

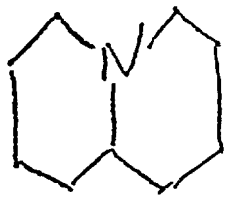
where xn is the temporary label, or Zm for a permanent one.

.22 ORTHOMESH

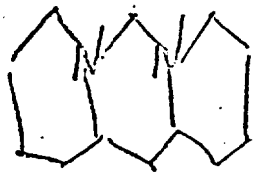
The Orthomesh consists of the VERTEX GROUP, VG, and the EDGE COUNT LIST, ECL. Together these elements define a planar mesh whose shape has become conventionalized by chemical usage - it corresponds to a carbocyclic ring system on which changes from C to other atoms can be mapped by exception. Examples which anticipate later notation follow:

3.221

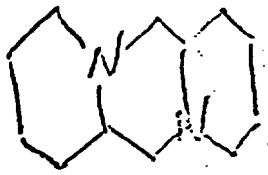
VG ECL VM (canonical and automorphisms)



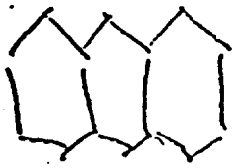
2A (1) (0,4,4) N(1)



4AA (0,1,0,1,4,4) N2(1,4) (2,3)

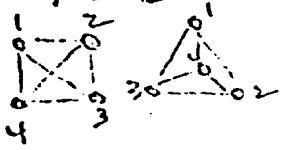


4ABA same N2(1,3) (2,4)

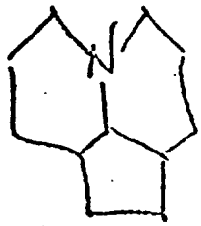


4BA same — —

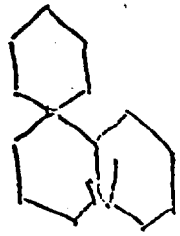
4BB (0,0,2,3,3,0) N(2)



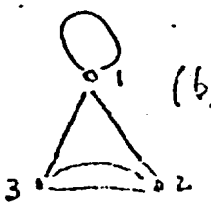
Note how the ECL fixes the symmetry!



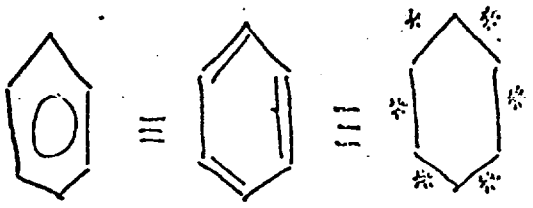
4BBB same N(1)



(6,0,0,3,4) N(3)



In 3.2 1, the VG is (0), i.e., the single ring, sans vertices, on which is mapped a single path.



The inscribed O stands for the aromatic character of the ring, i.e., a path with alternating single and double bonds; C:C.C:C.C:C.

Chemical evidence shows that in such a case, each bond is equivalent and has a character intermediate between "." and ":". We indicate this by the notation * which might be denoted in a radical as *C*C*C*C*C*C* or implicitly as 6*.

More complex rings have 2(r-1) vertices, where r is the number of included ring units (inscribed faces) as conventionally counted by chemists. Dendral II elaborates all possible vertex groups. The majority of organic ringed molecules fall in the category VG=(0), and $r \leq 4$, $v \leq 6$ covers as large a scope as the unpecialized generator is likely to be able to handle. These are reproduced here for convenience as Figure 3.21 and 3.22.

The v vertices implied by the VG must now be specified, and then the 3v/2 edges whose lengths are given by the ECL.

3.2.3 VERTEX MODIFIERS, VM.

The EDGE COUNT LIST, ECL, has a higher priority than the vertex modifiers, VM, list, an implied list of the vertex atoms which specifies (a) their composition and (b) their chirality. Since the most usual situation is all carbons we locate exceptions to that standard.

The VM list and its value is the 3-ple (vector)

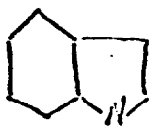
1) the non-carbon atoms in DENDRAL order, expressed as a composition, e.g., N2, or NIL.
2) A vertex locant list for these atoms.
3) An explicit list of the chirality of successive vertex atoms - (".", "+", "-", "A") referring to "unspecified", "dextro", "levo", "racemic"; see section 3.4 for details.

At any place, the omission of a list or the value NIL, implies "unspecified".

Tables of symmetries to elicit canonical forms will be provided or can be deduced from the VG (see Figure 3.1). For the moment the simplest algorithm will be to try all the automorphisms exhaustively, and save only the canonical representations.

3.24 PATH LIST.

[The ring definition is now completed by mapping linear paths onto the edges of the VG. In taking account of the symmetries, the sense of the path must be considered. Several representations are possible; in previous specifications, we displayed a vector of all paths in edge-number sequence: thus (.) (.C.C.N.) (.C.C.C.C.) for



1 2 3 4 5 6 7

3.241

Each edge of a path list was mapped with a linear, two-ended radical, presented in previously defined notation]

The ring atoms are all enumerable; first the vertices, V1, V2, etc., then the path atoms in the order of their presentation in the formula.

We now find it more convenient and conformable to chemical notation to map rings by exception. The path list will then consist of 3 parts:

3.242 ECL (already in orthomesh)

1) A vector that assigns a count to <u>each edge in sequence</u>. Each node is implicitly numbered by its sequence in the formula

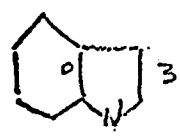
3.243 PATH MODIFIERS PM

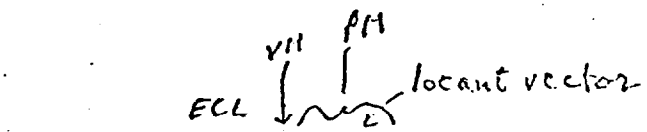
2) A composition indicating atoms replacing the implied C's.

This includes double bonds as U's. (abbreviations are available for aromaticity)

3.244 3) A locant vector for these replacing atoms and bonds

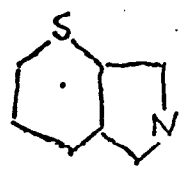
Example 3.241 then becomes:

2451 

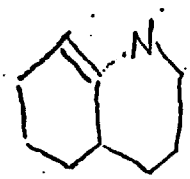


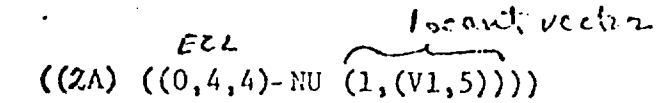
((2A) ((0,3,4)-N (1)))

orthomesh is 2A (0,3,4).

2452 

((2A) ((0,3,4)-NS (2,4)))

2453 



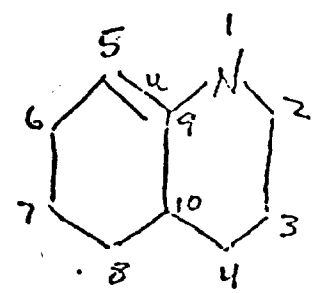
((2A) ((0,4,4)-NU (1,(V1,5))))

Caution: in internal notation all bonds are **numbered** as node-pairs. When the pair is (n,n+1) **only** the lower node need be numbered in external **notation.**

Note: in man-machine-interaction displays, parentheses can be replaced by formatted labels and indentations to facilitate editing and commands. We can visualize a display

2454

VG	2AA
VM	-
PL ECL	(0,4,4)
PLM	NU
N	1
U	(V1,5) or (5,9)

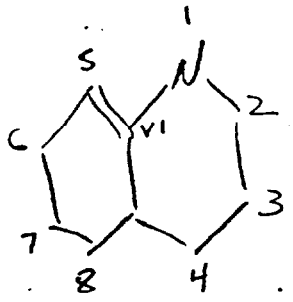


Vertex numbers can be replaced by arbitrary integers, e.g. numbered after all the path atoms. In this example, V1←9, V2←10.

By giving the edge count list a higher priority than the VM, We facilitate the storing of common orthomeshes under a familiar ring label. This can then be edited there for specific compounds. Thus, 3.2453can be input as

DECALIN

orthomesh label = VB,ECL PM
(DECALIN NU (1,(V1,5)))

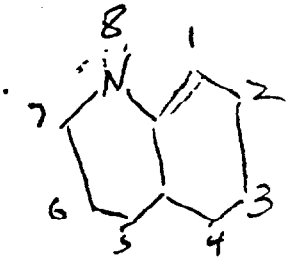


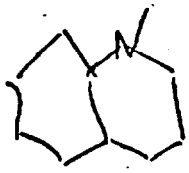
The chemists name for this is

Δ(1), 8-aza-decalin,

since he follows a somewhat different numbering system, (It should not be difficult to construct algorithms for interconverting many DENDRAL names with conventional chemical notation.

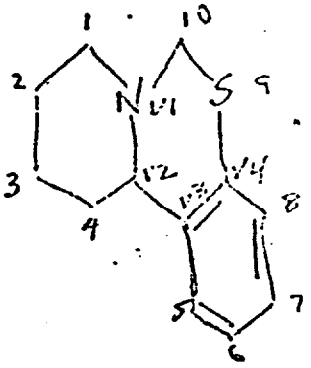
which gives the double bond priority.





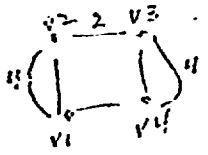
would, however, be 1-aza-decalin., closer to Dendral's (Decalin N (1))

Example of a complete definition with canonical numbering:



((4AA) (0,0,0,2,4,4) (N,1) (SU3 (10,v3,5,7)))

VG ECL VM PLM

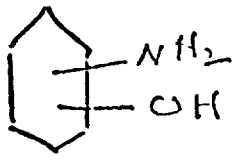


WHEN the aromatic ring is detected, this reduces to

((4AA) (0,0,*,2,4,4*) (N,1) (S,10))

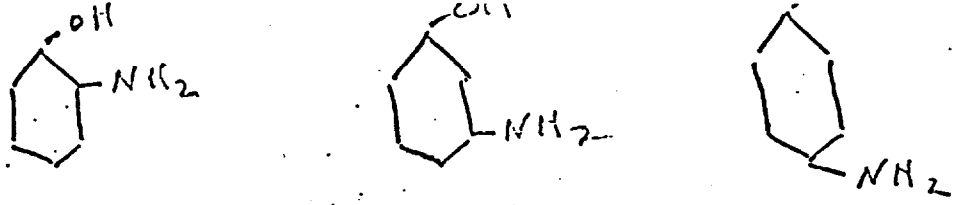
240 Embedding the ring in a tree.

The ring as now defined may be regarded as a superatom. However, it will, in general, have its own symmetries. Different locations have to be indicated for attaching further radicals. For example, the three amino-phenols are all special cases of



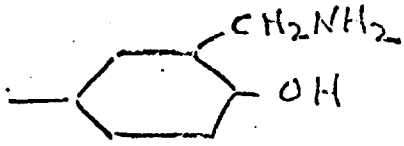
Z..NH2 OH (or Z..NO)

We will, then first list all of the attached radicals in DENDRAL order without reference to position of attachment. Then we will write a vector of locants describing the successive positions of attachment. For example,



Z(1,2)..NH2 OH Z(1,3)..NH2 OH Z(1,4)..NH2 OH

and a radical example



.Z(1,4,3)..OH CH2NH2

There are, of course, many automorphisms for the locant vector. In this example the locant vector is (1,4,3):

Canonical	Automorphisms
.Z(1,4,3)..OC.N	.Z(1,3,4)..C.N O
	.Z(4,1,2)..OC.N
	.Z(3,6,5)..OC.N
	etc.

The canonical form

1) lists the radicals in DENDRAL order, the afferent link being implied as first, and unsubstituted H's being ignored.

2) among the automorphic permutations of the locants, selects the least vector.

Resolving the ambiguities of the locant vectors is again a messy prospect and the simpliest general solution is to test all the symmetry operations seriatim and compare the locant vectors. More efficient rules will soon emerge for the frequent cases. However, as one suggestion, the pendant radicals might be replaced, during computation, by ordinal numbers in order to simplify comparisons.

When a - CH_2 - appears in a ring, two substitutions are possible at the same path atom. For the moment they are not distinguished. Chirality at vertices and path nodes is discussed later, 3.4.

CHIRALITY

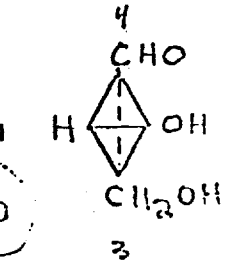
.410 For many purposes we emphasize the topological and can overlook the 3-dimensional spatial aspects of molecular structure. In real molecules, the lengths and rotations of bonds are understood to a widely varying precision, and may be greatly influenced by the history, energy-state, and immediate context of the molecule. In this sense, the topological connectivity is only a formal representation of a genus of states which can sometimes be inferred from it.

However, one aspect of stereo-chemistry generates distinctions among chemically stable species of very great importance, especially for biologically interesting compounds. This is chirality, which rests upon the alternating symmetry groups of the valence bonds of the C atom.

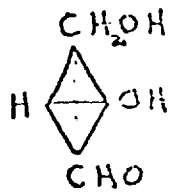
.411 The symmetries of the valences of carbon. The topological tetrahedron has the symmetry S4, that is all 4!= 24 permutations of its vertices (table 3.41). The carbon atom behaves, however, as a steric tetrahedron, so that two enantiomorphic, mirror images, must be distinguished for (C...abcd). These corresponding to odd and even alternating groups (table 3.410).

We can now regard the chirality of asymmetric carbons as a special case of the locant lists which specify substitutions on superatoms. The D- and L- isomers of, say, glyceraldehyde would be described (with implicit H's given leading locants, contra ring locants)

C(1,2,3,4) ....
H OH CH2.OH CHO



D-glyceraldehyde



L-

C(1,2,4,3)...
etc.

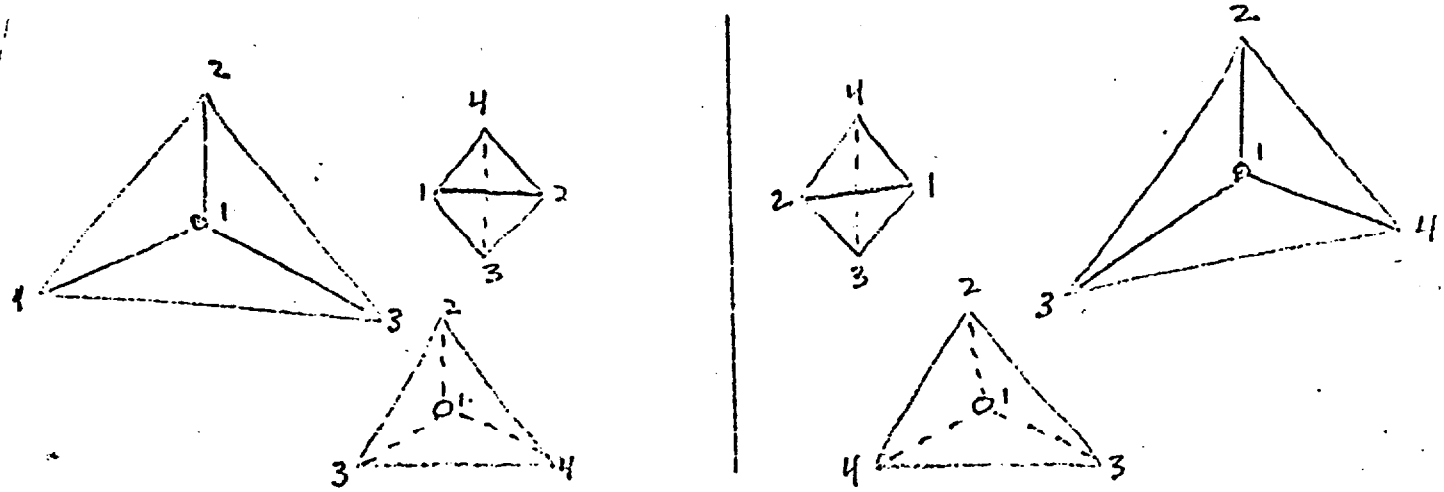
In view of the bisection of S_4 into the two A_4's, even and odd respectively, each locant list has 12 automorphisms (table 3.401). We can abbreviate C(1,2,3,4) as C(0) or "even", and C(1,2,4,3) as C(1) or "odd" respectively. We do not use the terms D-,L- or R-, S-, as these are based on a different set of rules for ordering the radicals (cf. Eliel, 1962; Prelog, 1966). However, the absolute configuration is fully defined in the present system, which lists the radicals in DENDRAL-hierarchical order. It is easy to program a translation between these systems, but DENDRAL is preferred for radical-generation. The rules, which are specially adapted for asymmetric carbon, are:

3.411 1) Radicals in DENDRAL sequence; however,

3.412 2) The afferent link, if any, takes position 4

3.413 3) Implied hydrogens are assigned implicit, leading locants

3.414 4) D-glyceraldehyde is taken as the prototype of the even group.



TETRAHEDRA

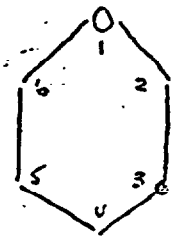
EVEN and ODD

C(0) C(1)

It is elementary that C(0)≠C(1) only when the C is asymmetric, i.e., the four substituents are all different. The generator program needs know only this. However, translation to D/L and R/S notation, and conservation of chirality may be important in analytical manipulations and in programming stereospecific reactions.

3.42 PATH ATOM ATTACHMENTS

Each H-atom of the path atoms of the pure ring is a candidate for replacement by a radical. If the path atom is a saturated C, $(-CH_2-)$, the two H's are not necessarily equivalent in chirality i.e., when the image of the rest of the ring differs, as seen from the C atom, via the afferent and efferant bonds. The image is obtained by cutting the bond, the attachment to the C is replaced by an H; the other cut edge then leads to a radical which can be evaluated by DENDRAL rules. Thus in



at (3), the image is (C.O.C.C.C.) afferently and C.C.C.O.C. efferently. Hence, any difference in the further substituents at (3) will make that atom asymmetric.

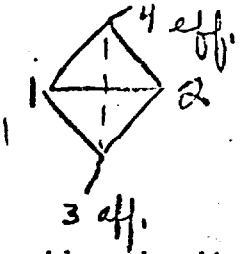
Customary notation fails to unify the several aspects of chirality: asymmetric C atoms in acyclic molecules, ring vertices and ring path atoms. DENDRAL furnishes (1) a systematic evaluation of radicals and (2) a convention for absolute configurattions by the allocation of the set

of radicals to the even or odd group, respectively.

At a trivalent ring vertex, the weights at the C are ordered by the canonical numbering of the VG edges defined by the orthomesh and further substituents (Paragraph 3.22-3.24) in the ring.

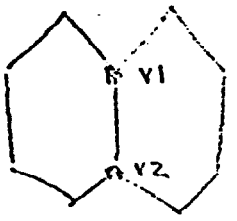
The chirality of the vertices of a ring is designated by a chirality status vector displayed in the VM list. This will be a string of 0 and 1 bits. Other characters may be used to designate "unspecified" and racemic conditions (statistical mixtures). "C(0)." can be economically replaced by "C+", "C(1)." by "C-".

For a path atom, a the afferent and efferent paths in the ring count as heaviest bond; Further substituents are then weighed as in Paragraph 3.41.

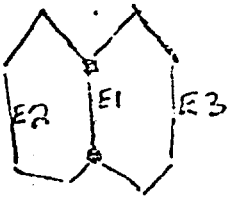


3.43 CIS-TRANS ISOMERISM

DENDRAL notation overrides the distinctions of CIS- and trans- configurations of adjacent vertices in rings, which appears in customary notation. In molecules of sufficient symmetry, e.g. decalin we will find an ambiguity: CIS will correspond to the chirality status vector (10) as well as (01); trans to (11) or (00), as we shall see:

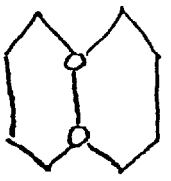


; the edges are numbered

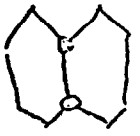


chirality at V1: V1(1,2,3,4) .... H E1 E2 E3 in canonical order. ~~These~~ V2 will be V2 (1,2,4,3) .... H E1 E2 E3. We thus have V1(0), V2(1) or the status vector (01).

The symmetry, however, also permits the molecule to be inverted, i.e., to



which gives the vector (10). The former is however canonical. By a corresponding argument, trans-decalin is



(00) or (11), the former canonical.

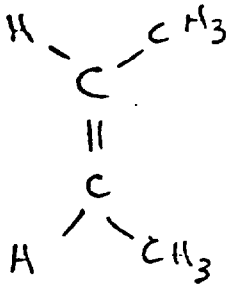
Note that the automorphisms of vertex chirality are resolved, and the path numbering fixed, before locants are established for path modifiers and appended radicals.

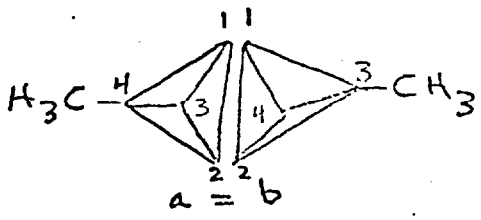
CIS/TRANS Isomerism at double bonds.

(Revision of 1 . 72)

The same approach of producing an absolute valence assignments can be used to describe geometrical isomerism at C = C.

The double bond is taken to occupy two adjacent positions of the first C atom encountered, postulated as having chirality status C(0). The second is then matched to the first. It will be C(0) or C(1) in an absolute, canonical description, and only one of these if the conditions for geometric isomerism are met. e.g.





cis-2-butene

: C (1,2,3,4).. H CH3 (a) C (1,2,4,3).. H CH3 (b)

odd

or : C.CH3 C(1). CH3 as an unambiguous description.

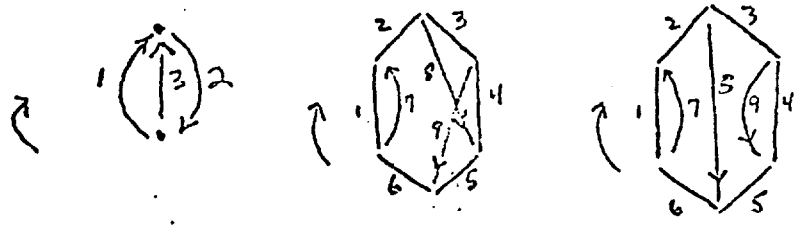
AXIAL AND EQUATORIAL

Hexagonal rings often adopt a chain-like configuration in which one of the H's at a path $-CH_2-$ may be found to lie either close to the plane of the ring (equatorial), the other above or below it (axial). These labels are part of a steric rather than topological description, but can sometimes be referred from the absolute configuration.

3.246 Revision of edge-numbering canons for twin edges.

(Revision of 2.34)

According to 2.34, the edges of a vertex group are numbered by circuiting the polygon (Hamilton circuit), than the chords as first encountered. Hence the numbering and polarity of these examples.



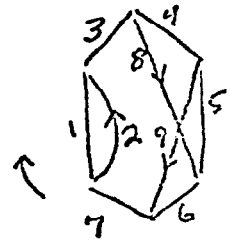
(2A) (6ABB) (6ACA)

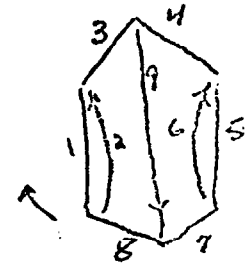
2A-1,2 6ABB-1 and -7, etc.

Edges -2 and -3, and -7 are twin edges, i.e., correspond to a span of 1.

The elaboration of symmetries is simplified if twin edges are not treated as chords, but are brought together in the sequence. Hence we now draw

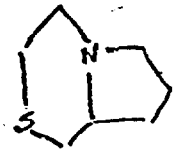
2A



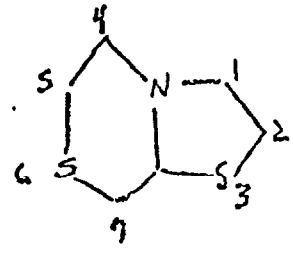


6ABB

The revision of edge order may influence the edge count list, canonical orientation of the orthomesh, and node-numbering in certain situations. For example



now becomes



in place of

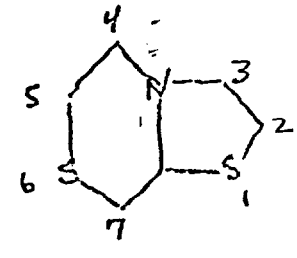


Table 3.1

CANONS OF DENDRAL ORDER (modified from DENDRAL [1])

Hierarchy of Vector Valuation in Decreasing Order of Significance

The DENDRAL-VALUE of a radical is a vector comprising:

R-COUNT

- Rings
- Other atoms (except H)

COMPOSITION OF RADICAL

- Rings
  - Composition CNOPSU*
  - Orthomesh
    - Vertex Group VG
    - Edge Count List ECL
  - Vertex Modifiers VM ; chirality status
  - Path modifiers PM
- Other atoms by atomic number

  C, N, O, P, S
  U (unsaturation is counted as a lowest-valued atom)

APICAL NODE

- Degree: Number of efferent radicals
- Composition of node:ring (by value), C,N,O,P,S
- Afferent link: (:, :, .)

APPENDANT RADICALS - attached to apical node

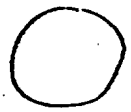
- vector of radicals in order of ascending* value
- locant list (ring) or chirality (atom) on apical node

---

* The sequence might be reversed

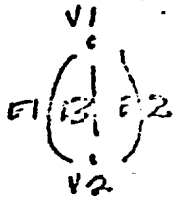
** to conform to chemists' notation which tends to assign higher-valued substituents to lower-numbered atoms, and orders atoms by valence as well as atomic number.

Fig 3.1 To illustrate permutation groups



1 path
no vertices

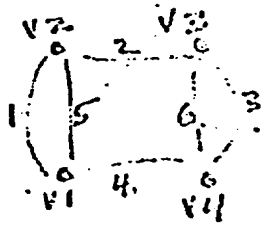
VG = 0



Symmetries

Nodes	Edges
12	1 2 3
21	1 3 2
	2 3 1
	2 1 3
	3 1 2
	3 2 1

VG = 2A



1234	1 2 3 4 5 6
3412	5 2 3 4 1 6
2143	1 2 6 4 5 3
4321	5 2 6 4 1 3
	6 4 5 2 3 1
	6 2 5 4 3 1
	6 2 1 4 3 5
	6 4 1 2 3 5
	1 4 6 2 5 3
	1 4 3 2 5 6

VG = 4AA

etc. 16 permutations

TABLE 3.410

The alternating groups A_4 (even) and (odd)

even permutation	C(0) cycles	odd permutation	C(1) cycles
1234	(1)(2)(3)(4)	1243	(1)(2)(34)
1342	(1)(234)	1324	(1)(23)(4)
1423	(1)(243)	1432	(1)(24)(3)
2143	(12)(34)	2134	(12)(3)(4)
3214	(123)(4)	2341	(1234)
2431	(124)(3)	2413	(1243)
3124	(132)(4)	3142	(1342)
3412	(13)(24)	3421	(1324)
3241	(134)(2)	3214	(13)(2)(4)
4132	(142)(3)	4123	(1432)
4213	(143)(2)	4231	(14)(2)(3)
4321	(14)(23)	4312	(1423)

Together, these constitute the symmetric group, S_4.

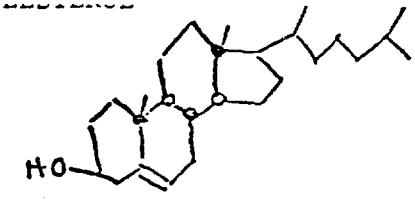
Example: 1234 ≈ 2143

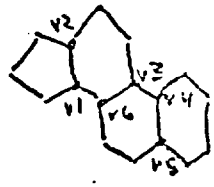
Ring-symmetries in DENDRAL.

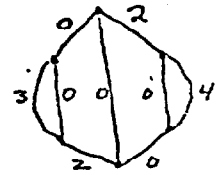
The input problem is to "produce all isomers of C10H14 containing at least one benzene ring." The verbose output displays the program's listing of the permutation group and other characteristics of the ring, before listing the isomers. Chirality is disregarded for this example (e.g. at #12).

```
(ISOMERS (QUOTE C10H14))
(MARCH-8-1968 VERSION)
C4*PHEN*H14
(RING *PHEN* COMPOSITION ((U . 4.) (C . 6.)) VALENCE (1. 1. 1. 1. 1.
1.) SYMMETRIES ((2. 3. 4. 5. 6. 1.) (3. 4. 5. 6. 1. 2.) (4. 5. 6. 1. 2
. 3.) (5. 6. 1. 2. 3. 4.) (6. 1. 2. 3. 4. 5.) (6. 5. 4. 3. 2. 1.) (5.
4 3. 2. 1. 6.) (4. 3. 2. 1. 6. 5.) (3. 2. 1. 6. 5. 4.) (2. 1. 6. 5. 4
. 3.) (1. 6. 5. 4. 3. 2.)) UNIQUE-NODES (1.))
MOLECULES       NO DOUBLE BOND EQUIVS
  1.    CH2..  C2H5   CH2.(*PHEN* 1.)H5 ,
  2.    CH2..  C2H5   (*PHEN* 1. 2.)H4.CH3 ,
  3.    CH2..  C2H5   (*PHEN* 1. 3.)H4.CH3 ,
  4.    CH2..  C2H5   (*PHEN* 1. 4.)H4.CH3 ,
  5.    (*PHEN* 1. 2.)H4..  C2H5   C2H5 ,
  6.    (*PHEN* 1. 3.)H4..  C2H5   C2H5 ,
  7.    (*PHEN* 1. 4.)H4..  C2H5   C2H5 ,
  8.    CH...  CH3   CH3   CH2.(*PHEN* 1.)H5 ,
  9.    CH...  CH3   CH3   (*PHEN* 1. 2.)H4.CH3 ,
 10.    CH...  CH3   CH3   (*PHEN* 1. 3.)H4.CH3 ,
 11.    CH...  CH3   CH3   (*PHEN* 1. 4.)H4.CH3 ,
 12.    CH...  CH3   (*PHEN* 1.)H5   C2H5 ,
 13.    (*PHEN* 1. 2. 3.)H3...  CH3   CH3   C2H5 ,
 14.    (*PHEN* 1. 2. 4.)H3...  CH3   CH3   C2H5 ,
 15.    (*PHEN* 1. 3. 2.)H3...  CH3   CH3   C2H5 ,
 16.    (*PHEN* 1. 3. 4.)H3...  CH3   CH3   C2H5 ,
 17.    (*PHEN* 1. 3. 5.)H3...  CH3   CH3   C2H5 ,
 18.    (*PHEN* 1. 4. 2.)H3...  CH3   CH3   C2H5 ,
 19.    C....  CH3   CH3  - CH3   (*PHEN* 1.)H5 ,
 20.    (*PHEN* 1. 2. 3. 4.)H2....  CH3   CH3   CH3   CH3 ,
 21.    (*PHEN* 1. 2. 3. 5.)H2....  CH3   CH3   CH3   CH3 ,
 22.    (*PHEN* 1. 2. 4. 5.)H2....  CH3   CH3   CH3   CH3 ,

DONE
```



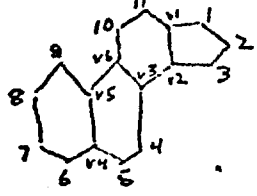




VG = (6ACA)

ECL = (0,3,0,2,0,4,0,2,0)

The canonical numbering details then become:



VM: No heteroatom substitutions; the chirality vector is given by:

v1....H 1 8 2 (1)	v4..=4 5 6 (3)	
v2....H 1 2 3 (0)	v5....H 5 6 7 (0)	(100300)
v3....H 3 4 9 (0)	v6....H 7 8 9 (0)	

(3) at v4 refers to the abrogation of chirality at a double-bonded vertex.

PLM: No heteroatoms, but a double bond at (v4,5).

The complete ring is then (x1 = nor-Δ^5-androstene)

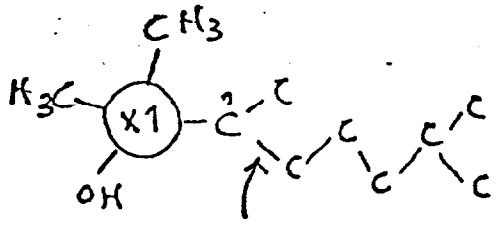
(x1 (6ACA (0,3,0,2,0,4,0,2,0)) (100300) (U ((v4,5)))

------orthomesh---------- --VM---- -----PLM------

VG ECL

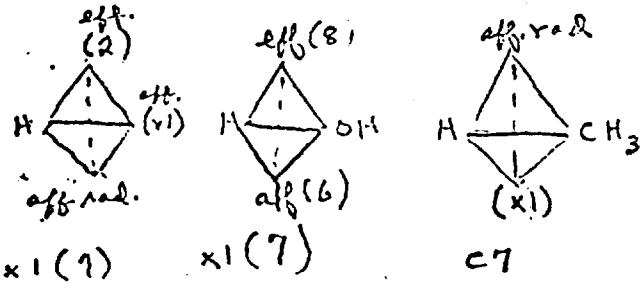
- - - - - - - - - - - - - - -

In the whole molecule, the canonical center is at↑ and we have:



.C3H6.CH..CH3 CH3 CH..CH3 x1(1,v1,v5,7)...CH3 CH3 OH
1-3 4 5 6 7 8 9 10 11 12

We have still to consider the chirality at x1(1) and x1(7) and C7. These prove to be



x1(1) x1(7) C7

rdl = - even = + even = +

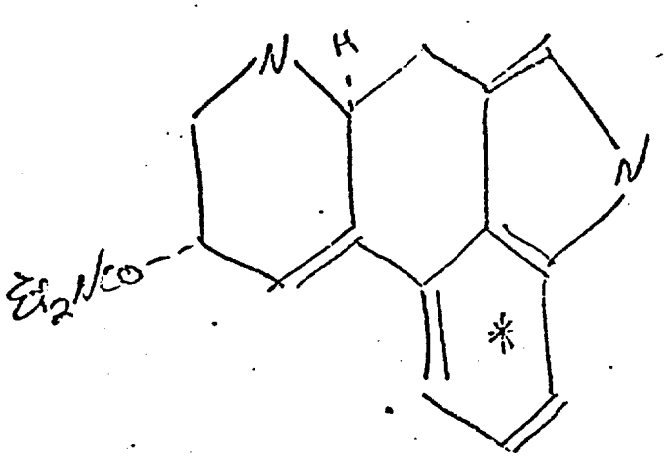
(v1)<(2)<eff.rad. H<OH<(6)<(8)

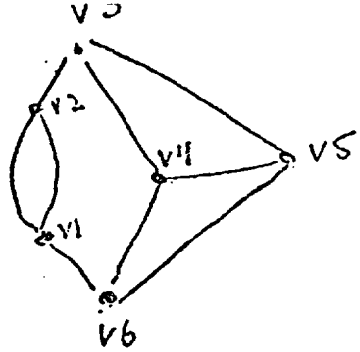
H<CH3<(x1)etc<eff.rad

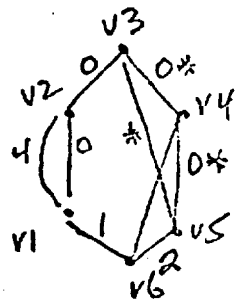
giving the completed formula:

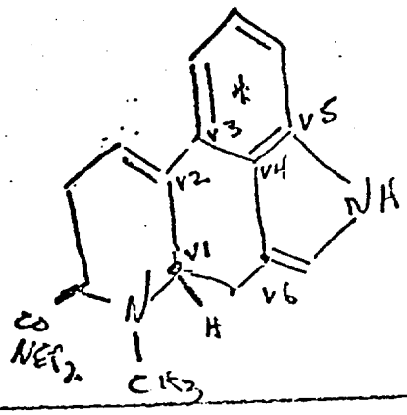
.C3H6.CH..CH3 CH3

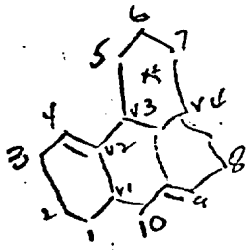
CH+.CH3 x1(1+,v1,v5,7-)...CH3 CH3 OH











canonical ordering algorithm by the ozkarovich

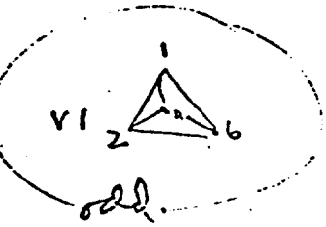
---

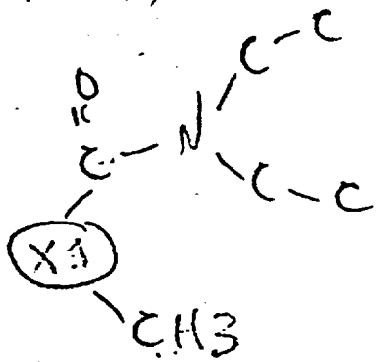
1 =

EZL = 0 4 0 0* 0* 2 1 0* 0

VM = 1 3 * * * 3

PM = (N2 U2 (1,8,4,9))





N...C2H5 C2H5 C:.O X1(1).CH3

Corrections (see 3.246

3.221 The vector for ECL of becomes

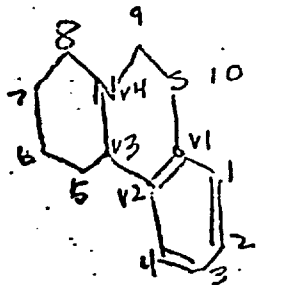
(0,4,1,0,4,1)

(5,0,0,4,3)

3.240 renumber to 3.3

The paragraph immediately preceding this should be numbered 3.2456

3.2456



new numbering

new ECL

(0,4,0,0,4,2)

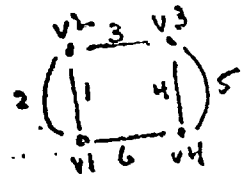
VH: (N,4)

PLM: (543 (10, V1, 1, 3))

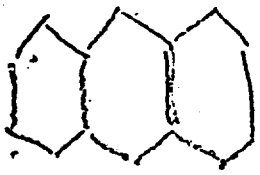
or, as an aromatic.

((4AA) (0*,4*,0,0,4,2) (N,4) (S' (10)))

Fig 3.1 4AA should be

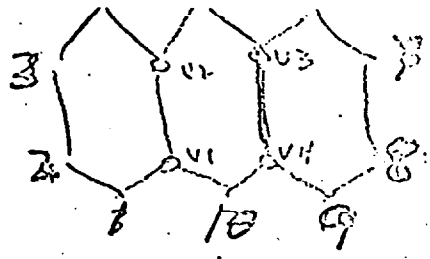


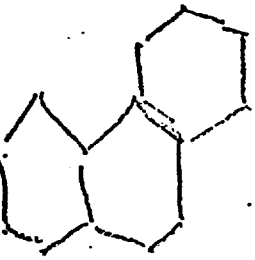
corresponding change in permutation table.



4AA

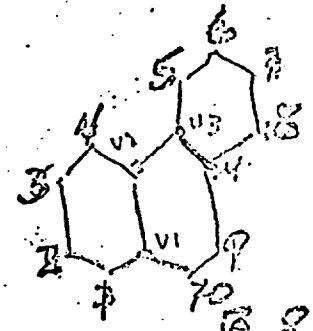
0,4,1,0,4,1

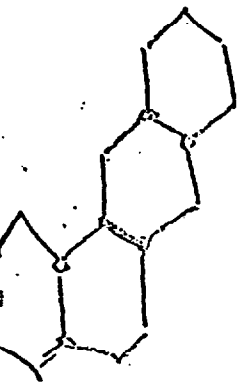


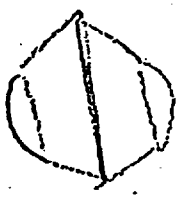


4BA

0,4,0,0,4,2

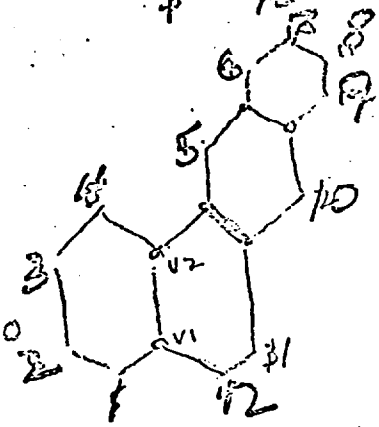






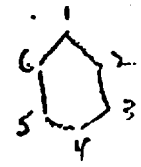
6ACA

0,4,0,1,0,4,1,2,0

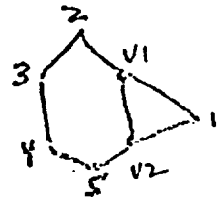


0 0

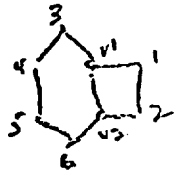
0 6 etc.



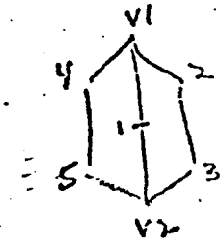
2A 0,1,4



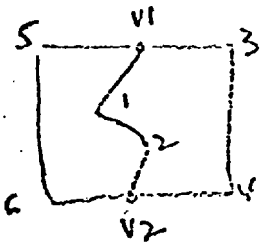
2A 0,2,4



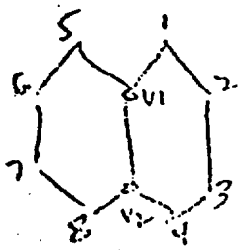
2A 1,2,2

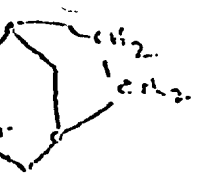


2A 2,2,2



2A 0,4,4





2A 1,2,3